Non-Uniform Attacks Against Pseudoentropy*

Krzysztof Pietrzak[†1] **and Maciej Skorski**[‡2]

1 **Institute of Science and Technology Austria, Klosterneuburg, Austria**
pietrzak@ist.ac.at
2 **Institute of Science and Technology Austria, Klosterneuburg, Austria**
mskorski@ist.ac.at

Abstract

De, Trevisan and Tulsiani [CRYPTO 2010] show that every distribution over n-bit strings which has constant statistical distance to uniform (e.g., the output of a pseudorandom generator mapping $n-1$ to n bit strings), can be distinguished from the uniform distribution with advantage ϵ by a circuit of size $O(2^n\epsilon^2)$.

We generalize this result, showing that a distribution which has less than k bits of min-entropy, can be distinguished from any distribution with k bits of δ-smooth min-entropy with advantage ϵ by a circuit of size $O(2^k\epsilon^2/\delta^2)$. As a special case, this implies that any distribution with support at most 2^k (e.g., the output of a pseudoentropy generator mapping k to n bit strings) can be distinguished from any given distribution with min-entropy $k+1$ with advantage ϵ by a circuit of size $O(2^k\epsilon^2)$.

Our result thus shows that pseudoentropy distributions face basically the same non-uniform attacks as pseudorandom distributions.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases pseudoentropy, non-uniform attacks

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.39

1 Introduction

De, Trevisan and Tulsiani [2] show a non-uniform attack against any pseudorandom generator (PRG) which maps $\{0,1\}^{n-1} \to \{0,1\}^n$. For any $\epsilon \geq 2^{-n/2}$, their attack achieves distinguishing advantage ϵ and can be realized by a circuit of size $O\left(2^n\epsilon^2\right)$. Their attack doesn't even need the PRG to be efficiently computable.

In this work we consider a more general question, where we ask for attacks distinguishing a distribution from any distribution with slightly higher min-entropy. We generalize [2], showing a non-uniform attack which, for any $\epsilon, \delta > 0$, distinguishes any distribution with $< k$ bits of min-entropy from any distribution with k bits of δ-smooth min-entropy with advantage ϵ, and where the distinguisher is of size $O(2^k\epsilon^2/\delta^2)$. As a corollary we recover the [2] result, showing that the output of any pseudoentropy generator $\{0,1\}^k \to \{0,1\}^n$ can be distinguished from any variable with min-entropy $k+1$ with advantage ϵ by circuits of size $O(2^k\epsilon^2)$.

- From a theoretical perspective, we prove where the separation between pseudoentropy and smooth min-entropy lies, by classifying how powerful computationally bounded adversaries can be so they can still be fooled to "see" more entropy than there really is.

* The full version is available at https://arxiv.org/abs/1704.08678
† Supported by the European Research Council, ERC consolidator grant (682815 – TOCNeT).
‡ Supported by the European Research Council, ERC consolidator grant (682815 – TOCNeT).

44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).
Editors: Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl;
Article No. 39; pp. 39:1–39:13


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

- From a more practical perspective, our result shows that using pseudoentropy instead of pseudorandomness (which for many applications is sufficient and allows for saving in entropy *quantity* [3]), will not give improvements in terms of *quality* (i.e., the size and advantage of distinguishers considered), at least not against generic non-uniform attacks.

1.1 Notation and Basic Definitions

Two variables X and Y are (s,ϵ) indistinguishable, denoted $X \sim_{s,\epsilon} Y$, if for all boolean circuits D of size $|D| \leq s$ we have $|\Pr[D(X)=1] - \Pr[D(Y)=1]| \leq \epsilon$. The statistical distance of X and Y is $d_1(X;Y) \stackrel{\mathsf{def}}{=} \sum_x |P_X(x) - P_Y(x)|$ (where $P_X(x) \stackrel{\mathsf{def}}{=} \Pr[X=x]$), the Euclidean distance of X and Y is $d_2(P_X;P_Y) \stackrel{\mathsf{def}}{=} \sqrt{\sum_x (P_X(x)-P_Y(x))^2}$. A variable X has min-entropy k if it doesn't take any particular outcome with probability greater 2^{-k}, it has δ-smooth min-entropy k [6], if it's δ close to some distribution with min-entropy k. X has k bits of HILL pseudoentoentry of quality (s,ϵ) if there exists a Y with min-entropy k that is (s,ϵ) indistinguishable from X, we use the following standard notation for these notions:

min-entropy: $\mathsf{H}_\infty(X) \stackrel{\mathsf{def}}{=} -\log \max_x (\Pr[X=x])$.

smooth min-entropy: $\mathsf{H}^\delta_\infty(X) \stackrel{\mathsf{def}}{=} \max_{Y, d_1(X;Y)\leq\delta} \mathsf{H}_\infty(Y)$.

HILL pseudoentropy: $\mathsf{H}^{\mathsf{HILL}}_{s,\epsilon}(X) \stackrel{\mathsf{def}}{=} \max_{Y, Y\sim_{(s,\epsilon)}X} \mathsf{H}_\infty(Y)$.

1.2 Our Contribution

In this work give generic non-uniform attacks on pseudoentropy distributions. A seemingly natural goal is to consider a distribution X with $\mathsf{H}_\infty(X) \leq k$ bits of min-entropy, strictly larger $\mathsf{H}^{\mathsf{HILL}}_{s,\epsilon}(X) \geq k+1$ bits of HILL entropy, and then give an upper bound on s in terms of ϵ. This does not work as there are X where $\mathsf{H}_\infty(X) \ll \mathsf{H}^\delta_\infty(X)$,[1] and as by definition $\mathsf{H}^\delta_\infty(X) = \mathsf{H}^{\mathsf{HILL}}_{\infty,\delta}(X)$, we can have a large entropy gap $\mathsf{H}^{\mathsf{HILL}}_{\infty,\delta}(X) - \mathsf{H}_\infty(X)$ even when considering unbounded adversaries against HILL entropy. For this reason, in our main technical result 1 below, we must consider distributions with bounded *smooth* min-entropy. This makes the statement of the lemma somewhat technical. In practice, the distributions considered often have bounded support, for example because they were generated from a short seed by a deterministic process (like a pseudorandom generator). In this case we can drop the smoothness requirement as stated in Theorem 2 below.

▶ **Lemma 1** (Nonuniform attacks against pseudoentropy). *Suppose that $X \in \{0,1\}^n$ does not have k bits of δ-smooth min-entropy, i.e., $\mathsf{H}^\delta_\infty(X) < k$, then for any ϵ we have*

$$\mathsf{H}^{\mathsf{HILL}}_{\tilde{O}(2^k\epsilon^2\delta^{-2}),\epsilon}(X) < k$$

where $\tilde{O}(\cdot)$ hides a factor linear in n.

▶ **Theorem 2.** *Let $f : \{0,1\}^k \to \{0,1\}^n$ be a deterministic (not necessarily efficient) function. Then we have*

$$\mathsf{H}^{\mathsf{HILL}}_{\tilde{O}(2^k\epsilon^2),\epsilon}(f(U_k)) \leq k+1.$$

more generally, for any X over $\{0,1\}^n$ with support of size $\leq 2^k$

$$\mathsf{H}^{\mathsf{HILL}}_{\tilde{O}(2^k\epsilon^2),\epsilon}(X) \leq k+1.$$

[1] Consider an X which is basically uniform over $\{0,1\}^n$, but has mass δ on one particular point, then $\log \delta^{-1} = \mathsf{H}_\infty(X) \ll \mathsf{H}^\delta_\infty(X) = n$.

▶ Remark (Concluding best attacks against PRGs). For the special case $n = k + 1$ we recover the bound for pseudo*random* generators from [2].

Proof of Theorem 2. The theorem follows from Lemma 1 when $\delta = 1/2$; consider any X with support of size $\leq 2^k$, then $\mathsf{H}^{\delta}_{\infty}(X) \leq k + 1$, as no matter how we cut probability mass of $1 - \delta = 1/2$ over 2^k elements, one element will have the weight at least 2^{-k-1}. ◀

1.3 Proof Outline

1.3.1 A Weaker Result as a Ball-Bins Problem

We outline the proof of a somewhat weakened version of Theorem 2 in the language of balls and bins. For every Y of min-entropy $k' = k + \Omega(1)$ we want to distinguish Y from $X = f(U_k)$. Suppose for simplicity that Y is flat and f is injective, so that X is also flat. Our strategy will be to hash the points randomly into two bins and take advantage of the fact that the *average maximum load* is closer to $\frac{1}{2}$ when we sample from Y than when drawing from X. The reason is that Y has more balls, so by the law of large numbers, we expect the load to be "more concentrated" around the mean.

Think of throwing balls (inputs x) into two bins (labeled by -1 and 1). If the balls come from the support of X, the expected maximum load (over two bins) equals $\approx 2^{k-1} + \sqrt{2/\pi} \cdot 2^{k/2}$. Similarly, if the balls come from the support of Y, then maximum load is $2^{k'-1} + \sqrt{2/\pi} \cdot 2^{k'/2}$. In terms of the average load (the load normalized by the total number of balls):

$$\mathsf{AverageMaxLoad}(X) \approx 0.5 + \sqrt{2/\pi} \cdot 2^{-k/2} \quad \text{w.h.p. when drawing from } X\,,$$
$$\mathsf{AverageMaxLoad}(Y) \approx 0.5 + \sqrt{2/\pi} \cdot 2^{-k'/2} \quad \text{w.h.p. when drawing from } Y\,.$$

As $k' = k + \Omega(1)$ we obtain (with good probability):

$$\mathsf{AverageMaxLoad}(X) - \mathsf{AverageMaxLoad}(Y) = \Omega(2^{-k/2}).$$

Letting D be one of these bins assignments we obtain a distinguisher with advantage $\epsilon = \Omega(2^{-k/2})$. To generate the assignments efficiently we relax the assumption about choosing bins and assume only that the choices of bins are independent for any group of $\ell = 4$ balls. The fourth moment method allows us to keep sufficiently good probabilistic guarantees on the maximum load.

1.3.2 The General Case by Random Walk Techniques

1.3.2.1 A high-level outline and comparison to [2]

Below in Figure 1 we sketch the flow of our argument.

Our starting point is the proof from [2]. They use the fact that a random mapping $\mathsf{D} : \{0,1\}^n \to \{-1,1\}$ likely distinguishes any two distributions X and Y over $\{0,1\}^n$ with advantage being the Euclidean distance $d_2(X;Y) \stackrel{\mathsf{def}}{=} \sqrt{\sum_x (P_X(x) - P_Y(x))^2}$.

For any X and Y with constant statistical distance $\sum_x |P_X(x) - P_Y(x)| = \Theta(1)$ (which is the case for the PRG setting where $Y = U_n$ and $X = PRG(U_{n-1})$) this yields a bound $\Omega\left(2^{-\frac{n}{2}}\right)$. This bound can be then amplified, at the cost of extra advice, by partitioning the domain $\{0,1\}^n$ and combining corresponding advantages (advice basically encodes if there is a need for flipping the output). Finally one can show that 4-wise independence provides enough randomness for this argument, which makes sampling D efficient. Our argument deviates from this approach in two important aspects.

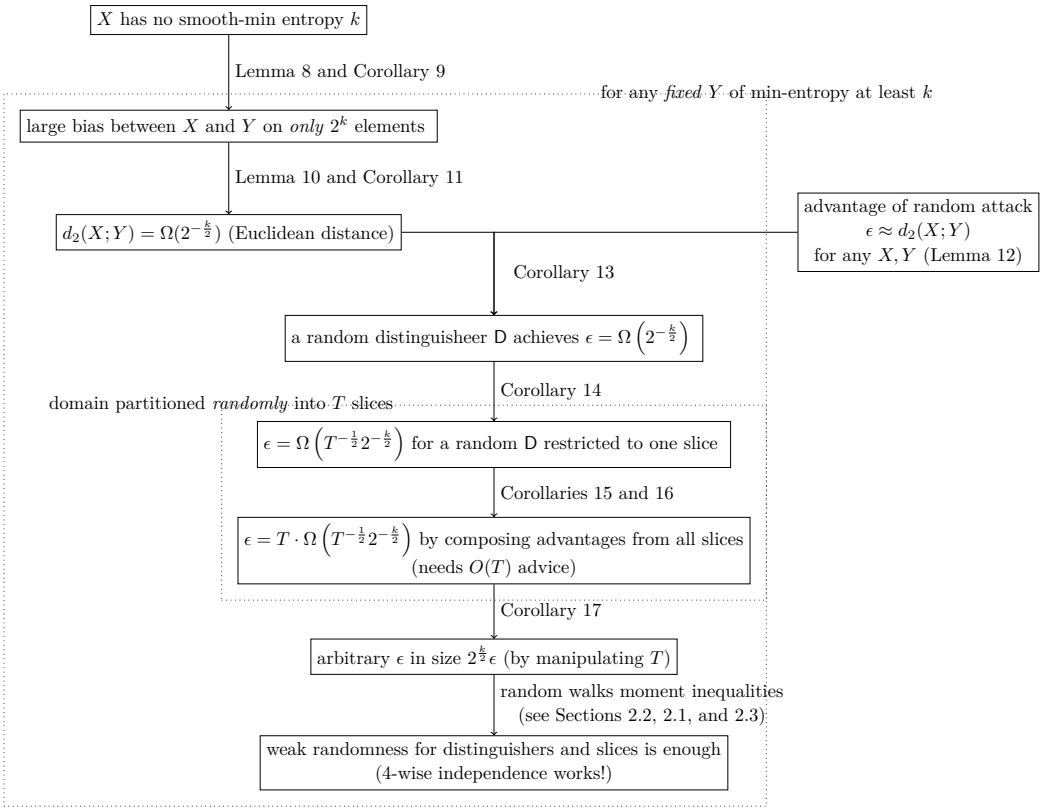


Figure 1 The map of our proof.

The first difference is that in the pseudoentropy case we can improve the advantage from $\Omega\left(2^{-\frac{n}{2}}\right)$, where n is the logarithm of the support of the variables considered, to $\Omega\left(2^{-\frac{k}{2}}\right)$, where k is the min-entropy of the variable we want to distinguish from. The reason is that being statistically far from any k-bit min-entropy distributions implies a *large bias on already* 2^k *elements*. This fact (see Lemma 8 and Corollary 9, and also Figure 3) is a new characterization of smooth min-entropy of independent interest.

The second subtlety arises when it comes to amplify the advantage over the partition slices. For the pseudorandomness case it is enough to split the domain in a deterministic way, for example by fixing prefixes of n-bit strings, in our case this is not sufficient. For us a "good" partition must shatter the 2^k-element high-biased set, which can be arbitrary. Our solution is to use *random partitions*, in fact, we show that using 4-universal hashing is sufficient. Generating base distinguishers and partitions at the same time makes probability calculations more involved.

Technical calculations are based on the fourth moment method, similarly as in [2]. The basic idea is that for settings where the second and fourth moment are easy to compute (e.g. sums of independent symmetric random variables) we can obtain good upper and lower bounds on the first moment. In the context of algorithmic applications these techniques are usually credited to [1]. Interestingly, exploiting natural relations to *random walks*, we show that calculations immediately follow by adopting classical (almost one century old) tools and results [5, 4]. Our technical novelty is an application of moment inequalities due to Marcinkiewicz-Zygmund and Paley-Zygmund, which allow us to prove slightly more than just the existence of an attack. Namely we generate it with constant success probability.

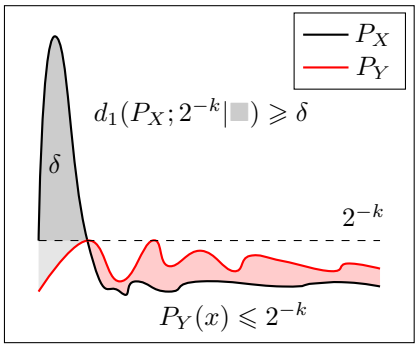


Figure 2 An intuition behind the attack. Random ± 1-weights make the bias equal to the ℓ_2-distance of P_X and P_Y. This distance can be bounded in terms of the ℓ_1 distance, which concentrates mass difference δ on less than 2^k elements (the region in gray).

1.3.2.2 Advantage $\Omega(2^{-k/2})$

Consider any X with δ-smooth min-entropy smaller than k. This requirement can be seen as a statement about the "shape" of the distribution. Namely, the mass of X that is above the threshold 2^{-k} equals at least δ, that is

$$\sum_x \max(P_X(x) - 2^{-k}, 0) \geqslant \delta.$$

For an illustration see Figure 2.

We construct our attack based on this observation. Define the advantage of a function D for distributions X and Y as

$$\mathsf{Adv}^{\mathsf{D}}(X;Y) = \left| \sum_x \mathsf{D}(x)(P_X(x) - P_Y(x)) \right|$$

(writing also $\mathsf{Adv}^{\mathsf{D}}_S$ when the summation is restricted to a subset S). Consider a random distinguisher $\mathsf{D} : \{0,1\}^n \to \{-1,1\}$. Random variables $\mathsf{D}(x)$ for different x are independent, have zero-mean and second moment equal to 1. Therefore the expected square of of the advantage, over the choice of D, equals

$$\mathbb{E}\left[\left(\mathsf{Adv}^{\mathsf{D}}(X;Y)\right)^2\right] = \mathbb{E}\left|\sum_x \mathsf{D}(x)(P_X(x) - P_Y(x))\right|^2 = \sum_x (P_X(x) - P_Y(x))^2 .$$

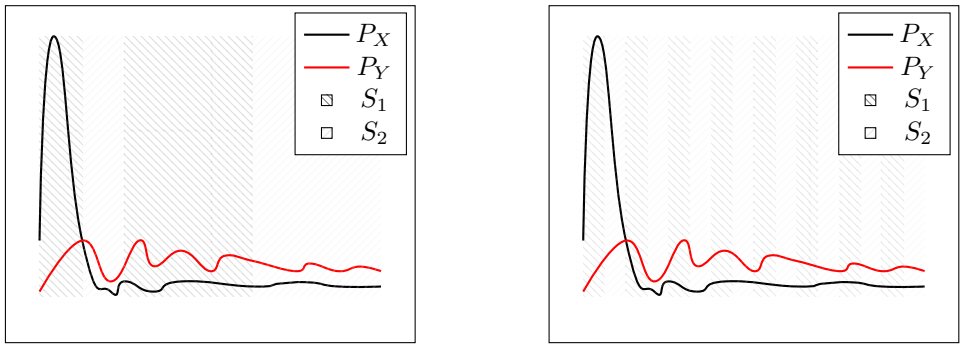
Let S be the set of x such that $P_X(x) > 2^{-k}$. For any Y of min-entropy at least k we obtain

$$\sum_{x\in S}(P_X(x) - P_Y(x))^2 \geqslant \sum_{x\in S}(P_X(x) - 2^{-k})^2 \geqslant |S|^{-1}\left(\sum_{x\in S}\left(P_X(x) - 2^{-k}\right)\right)^2 \geqslant 2^{-k}\delta^2$$

where the first inequality follows because $P_Y(x) \leqslant 2^{-k} < P_X(x)$ for $x \in S$, the second inequality is by the standard inequality between the first and second norm, and the third inequality follows because we showed that $\Pr[X \in S] \geqslant |S| \cdot 2^{-k} + \delta$ (illustrated in Figure 2) which also implies $|S|^{-1} \geqslant 2^{-k}$.

By the previous formula on the expected squared advantage this means that

$$\mathbb{E}\left[\left(\mathsf{Adv}^{\mathsf{D}}(X;Y)\right)^2\right] \geqslant 2^{-k}\delta^2$$



(a) An example of a "bad" partition. Almost all advantage is captured by one partition slice S_1.

(b) An example of a "good" partition. The advantage is evenly distributed among slices S_1, S_2.

Figure 3 Illustration of good and bad partitions.

for at least one choice of D. This implies

$$\mathsf{Adv}^{\mathsf{D}}(X;Y) \geqslant 2^{-\frac{k}{2}}\delta.$$

A random D as defined would be of size exponential in n, but since we used only the second moment in calculations, it suffices to generate $\mathsf{D}(x)$ as pairwise independent random variables. By assuming 4-wise independence – which can be computed by $O(n^2)$ size circuits – we can prove slightly more, namely that a constant fraction of generated D's are good distinguishers. This property will be important for the next step, where we amplify the advantage assuming larger distinguishers.

1.3.2.3 Leveraging the advantage by slicing the domain

Consider a random and equitable partition $\{S_i\}_{i=1}^{T}$ of the set $\{0,1\}^n$. From the previous analysis we know that a random distinguisher achieves advantage $\epsilon = d_2(P_X;P_Y)$ over the whole domain. Note that (for any, not necessarily random partition $\{S_i\}_i$) we have

$$\left(d_2(P_X;P_Y)\right)^2 = \sum_{i=1}^{T} \left(d_2(P_X;P_Y|S_i)\right)^2$$

where $d_2(P_X;P_Y|S_i)$ is the restriction of the distance to the set S_i (by restricting the summation to S_i). From a random partition we expect the mass difference between P_X and P_Y to be *distributed evenly* among the partition slices (see Figure 3(b)). Based on the last equation, we expect

$$d_2(P_X;P_Y|S_i) \approx \frac{d_2(P_X;P_Y)}{\sqrt{T}}$$

to hold with high probability over $\{S_i\}_i$.

In fact, if the mass difference is not well balanced amongst the slices (in the extreme case, concentrated on one slice) our argument will not offer any gain over the previous construction (see Figure 3(a)).

By applying the previous argument to individual slices, for every i we can obtain an advantage $\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \Omega\left((T^{-\frac{1}{2}}2^{-\frac{k}{2}})\delta\right)$ when restricted to the set S_i (with high probability

over the choice of D and $\{S_i\}_i$). Now if the sets S_i are *efficiently recognizable*, we can combine them into a better distinguisher. Namely for every i we chose a value $\beta_i \in \{-1,1\}$ such that D's advantage (before taking the absolute value) restricted to S_i has sign β_i, and set

$$\hat{\mathsf{D}}(x) = \beta_i \mathsf{D}(x), \text{ where } i \text{ is such that } x \in S_i,$$

then the advantage equals (with high probability over D and the S_i's)

$$\mathsf{Adv}^{\hat{\mathsf{D}}}(X;Y) = \sum_{i=1}^{T} \mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \Omega\left(T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta\right).$$

We need to specify a 4-wise independent hash for D, another 4-wise independent hash for deciding in which of the T slices an element lies, and T bits to encode the β_i's. Thus for a given T the size of $\hat{\mathsf{D}}$ will be $T + \tilde{O}(n)$. Using the above equation, we then get a smooth tradeoff $s = O(2^k \epsilon^2 \delta^{-2})$ between the advantage ϵ and the circuit size s. This discussion shows that to complete the argument we need the following two properties of the partition (a) the mass difference between P_X and P_Y is (roughly) equidistributed among slices and (b) the membership in partition slices can be efficiently decided.

1.3.2.4 Slicing using 4-wise independence

To complete the argument, we assume that T is a power of 2, and generate the slicing by using a 4-universal hash function $h : \{0,1\}^n \to \{0,1\}^{\log T}$. The i-th slice S_i is defined as $\{x \in \{0,1\}^n : h(x) = i\}$. These assumptions are enough to prove that

$$\mathbb{E}\, \mathsf{Adv}^{\hat{\mathsf{D}}}_{S_i}(X;Y) = \Omega\left(T^{-\frac{1}{2}} d_2(P_X;P_Y)\right) = \Omega\left(T^{-\frac{1}{2}} 2^{-\frac{k}{2}} \delta\right).$$

Interestingly, the expected advantage (left-hand side) cannot be computed directly. The trick here is to bound it in terms of the second and fourth moment. The above inequality, coupled with bounds on second moments of the advantage $\mathsf{Adv}^{\hat{\mathsf{D}}}_{S_i}$ (obtained directly), allows us to prove that

$$\Pr\left[\sum_{i=1}^{T} \mathsf{Adv}^{\hat{\mathsf{D}}}_{S_i} \geqslant \Omega(1) \cdot T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta\right] > \Omega(1).$$

This shows that there exists the claimed distinguisher $\hat{\mathsf{D}}$. In fact, a *constant fraction* of generated (over the choice of D and $\{\S_i\}_i$) distinguishers $\hat{\mathsf{D}}$'s works.

1.3.2.5 Random walks

From a technical point of view, our method involves computing higher moments of the advantages to obtain concentration and anti-concentration results. The key observation is that the advantage written down as

$$\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \left|\sum_x (P_X(x) - P_Y(x)) \mathbf{1}_{S_i}(x) \mathsf{D}(x)\right|$$

which can be then studied as a *random walk*

$$\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \left|\sum_x \xi_{i,x}\right|$$

with zero-mean increments $\xi_{i,x} = (P_X(x) - P_Y(x)) \mathbf{1}_{S_i}(x) \mathsf{D}(x)$. The difference with respect to classical model is that the increments are only ℓ-wise independent (for $\ell = 4$). However, that classical moment bounds still apply (see Sections 2.2 and 2.3 for more details).

2 Preliminaries

2.1 Interpolation Inequalities

Interpolation inequalities show how to bound the p-th moment of a random variable if we know bounds on one smaller and one higher moment. The following result is known also as *log-convexity of L_p norms*, and can be proved by the Hölder Inequality.

▶ **Lemma 3** (Moments interpolation). *For any $p_1 < p < p_2$ and any bounded random variable Z we have*

$$\|Z\|_p \leqslant (\|Z\|_{p_1})^{\theta} (\|Z\|_{p_2})^{1-\theta}$$

where θ is such that $\frac{\theta}{p_1} + \frac{1-\theta}{p_2} = \frac{1}{p}$, and for any r we define $\|Z\|_r = (\mathbb{E}\,|Z|^r)^{\frac{1}{r}}$.

Alternatively, we can *lower bound* a moment given *two higher moments*. This is very useful when higher moments are easier to compute. In this work will bound first moments from below when we know the second and the fourth moment (which are easier to compute as they are even-order moments)

▶ **Corollary 4.** *For any bounded Z we have $\mathbb{E}\,|Z| \geqslant \frac{(\mathbb{E}\,|Z|^2)^{\frac{3}{2}}}{(\mathbb{E}\,|Z|^4)^{\frac{1}{2}}}$.*

2.2 Moments of random walks

For a random walk $\sum_x \xi(x)$, where $\xi(x)$ are independent with zero-mean, we have good control over the moments, namely $\mathbb{E}\,|\sum_x \xi(x)|^p = \Theta(1) \cdot (\sum_x \mathrm{Var}(\xi(x)))^{\frac{p}{2}}$ where constants depend on p. This result is due to Marcinkiewicz and Zygmund [5] who extended the former result of Khintchine [4]. Below we notice that for *small moments* p it suffices to assume only p-wise independence (most often used versions assume fully independence)

▶ **Lemma 5** (Strengthening of Marcinkiewicz-Zygmund's Inequality for $p = 4$). *Suppose that $\{\xi(x)\}_{x\in\mathcal{X}}$ are 4-wise independent, with zero mean. Then we have*

$$\frac{1}{\sqrt{3}} \left(\sum_{x\in\mathcal{X}} \mathrm{Var}(\xi(x))\right)^{\frac{1}{2}} \leqslant \mathbb{E}\left|\sum_{x\in\mathcal{X}} \xi(x)\right| \leqslant \left(\sum_{x\in\mathcal{X}} \mathrm{Var}(\xi(x))\right)^{\frac{1}{2}},$$

$$\mathbb{E}\left|\sum_{x\in\mathcal{X}} \xi(x)\right|^2 = \sum_{x\in\mathcal{X}} \mathrm{Var}(\xi(x)),$$

$$\left(\sum_{x\in\mathcal{X}} \mathrm{Var}(\xi(x))\right)^2 \leqslant \mathbb{E}\left|\sum_{x\in\mathcal{X}} \xi(x)\right|^4 \leqslant 3\left(\sum_{x\in\mathcal{X}} \mathrm{Var}(\xi(x))\right)^2.$$

The proof appears in Section 4.1.

2.3 Anticontentration bounds

▶ **Lemma 6** (Paley-Zygmund Inequality). *For any positive random variable Z and a parameter $\theta \in (0,1)$ we have*

$$\Pr\left[Z > \theta\,\mathbb{E}\,Z\right] \geqslant (1-\theta)^2 \frac{(\mathbb{E}\,Z)^2}{\mathbb{E}\,Z^2}.$$

By applying Lemma 6 to the setting of Lemma 5, and choosing $\theta = \frac{1}{\sqrt{3}}$ we obtain:

▶ **Corollary 7** (Anticoncentration for walks with 4-wise independent increments). *Suppose that* $\{\xi(x)\}_{x\in\mathcal{X}}$ *are* 4*-wise independent with zero-mean, then we have*

$$\Pr\left[\left|\sum \xi(x)\right| > \frac{1}{3}\left(\sum \mathrm{Var}(\xi(x))\right)^{\frac{1}{2}}\right] > \frac{1}{17}.$$

where the summation is over $x \in \mathcal{X}$.

3 Proof of Lemma 1

▶ **Lemma 8** (Characterizing smooth min-entropy). *For any random variable* X *with values in a finite set* $\mathcal{X}$*, any* δ *and* k *we have the following equivalence*

$$H_\infty^\delta(X) \geqslant k \iff \sum_{x\in\mathcal{X}} \max\left(P_X(x) - 2^{-k}, 0\right) \leqslant \delta.$$

The proof appears in Section 4.2. We will work with the following equivalent statement

▶ **Corollary 9** (No smooth min-entropy k implies bias w.r.t. distributions of min-entropy k over at most 2^k elements). *We have* $\mathsf{H}_\infty^\delta(X) < k$ *if and only if there exists a set* S *of at most* 2^k *elements such that*

$$\sum_{x\in S} |P_X(x) - P_Y(x)| > \delta$$

for all Y *of min-entropy at least* k.

Proof of Corollary 9. The direction $\Longleftarrow$ trivially follows by the definition of smooth min-entropy. Now assume $\mathsf{H}_\infty^\delta(X) < k$. Let S be the set of all x such that $P_X(x) > 2^{-k}$, then $|S| < 2^k$, and moreover by Lemma 8 we have $\sum_{x\in S}\left(P_X(x) - 2^{-k}\right) > \delta$. In particular for any Y of min-entropy k (i.e., $P_Y(x) \leqslant 2^{-k}$ for all x)

$$\sum_{x\in S} (P_X(x) - P_Y(x)) > \delta\,. \qquad \blacktriangleleft$$

▶ **Lemma 10** (Bias implies Euclidean distance). *For any distributions* P_X, P_Y *on* $\mathcal{X}$ *and any subset* S *of* $\mathcal{X}$ *we have*

$$\left(\sum_{x\in S} (P_X(x) - P_Y(x))^2\right)^{\frac{1}{2}} > |S|^{-1/2} \sum_{x\in S} |P_X(x) - P_Y(x)|\,.$$

Proof. By the Jensen Inequality we have

$$|S|^{-1}\left(\sum_{x\in S} (P_X(x) - P_Y(x))^2\right) > \left(|S|^{-1}\sum_{x\in S} |P_X(x) - P_Y(x)|\right)^2$$

which is equivalent to the statement. ◀

▶ **Corollary 11** (No smooth min-entropy implies Euclidean distance to min-entropy distributions). *Suppose that* $\mathsf{H}_\infty^\delta(X) < k$*. Then for any* Y *of min-entropy at least* k *we have* $\left(\sum_x |P_X(x) - P_Y(x)|^2\right)^{\frac{1}{2}} > 2^{-\frac{k}{2}}\delta$.

Proof of Corollary 11. It suffices to combine Lemma 10 and Corollary 9. ◀

By Corollary 7 we conclude that the advantage of a random distinguisher for any two measures (in our case P_X and P_Y) equals the Euclidean distance.

▶ **Lemma 12** (The advantage of a random distinguisher equals the Euclidean distance). *Let $\{\mathsf{D}(x)\}_{x\in\{0,1\}^n}$ be 4-wise independent as indexed by x and such that $\mathsf{D}(x)$ outputs a random element from $\{-1,1\}$. Then for any set S we have*

$$\left|\sum_{x\in S}\mathsf{D}(x)(P_X(x)-P_Y(x))\right| > \frac{1}{3}\cdot d_2(P_X;P_Y)$$

with probability $\frac{1}{17}$ over the choice of D (the result actually holds for any measures in place of P_X, P_Y).

For our case, that is the setting in Lemma 10, we obtain

▶ **Corollary 13** (A random attack achieves $\Omega\left(2^{-k}\delta\right)$ with significant probability). *For X,Y as in Corollary 11, and D as in Lemma 12 we have $\mathsf{Adv}^{\mathsf{D}}(X;Y) \geq \frac{1}{3}\cdot 2^{-\frac{k}{2}}\delta$ w.p. $\frac{1}{17}$ over D.*

3.1 Partitioning the domain into T slices

Let $h : \{0,1\}^n \to [1\ldots 2^t]$, where $t = \lceil \log T \rceil$, be a 4-universal hash function. Define $S_i = \{x : h(x) = i\}$, $\Delta(x) = P_X(x) - P_Y(x)$ and consider advantages on slices S_i

$$\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \left|\sum_x \Delta(x)\mathsf{D}(x)\mathbf{1}_{S_i}(x)\right| .$$

The following corollary shows that on each of our T slices, we get the advantage $T^{-\frac{1}{2}}2^{-\frac{k}{2}}\delta$. The proof appears in Section 4.3.

▶ **Corollary 14** ((Mixed) moments of slice advantages). *For D, $\{S_u\}_u$ as above and every i,j*

$$\mathbb{E}_{\mathsf{D},\{S_u\}_u}\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) \geqslant 3^{-\frac{1}{2}}T^{-\frac{1}{2}}\cdot d_2\left(P_X;P_Y\right),$$
$$\mathbb{E}_{\mathsf{D},\{S_u\}_u}\left(\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y)\,\mathsf{Adv}^{\mathsf{D}}_{S_j}(X;Y)\right) \leqslant T^{-1}\cdot d_2(P_X;P_Y)^2,$$

(the statement is valid for arbitrary measures in place of P_X, P_Y).

Denote $Z = \sum_i \mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y)$. Using Lemma 6 with $\theta = \frac{1}{\sqrt{3}}$ where we compute $\mathbb{E}\,Z^2$ and $\mathbb{E}\,Z$ according to Corollary 14 we obtain $\Pr\left[|Z| > \frac{1}{\sqrt{3}}\cdot\mathbb{E}\,|Z|\right] \geqslant \frac{1}{17}$. Bounding once again $\mathbb{E}\,|Z|$ as in Corollary 14 we get

▶ **Corollary 15** (Total advantage on all partition slices). *For X,Y as in Corollary 11, D and S_i defined above we have*

$$\Pr_{\mathsf{D},\{S_u\}_u}\left[\sum_{i=1}^{T}\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) \geqslant \frac{1}{3}\cdot T^{\frac{1}{2}}2^{-\frac{k}{2}}\delta\right] \geqslant \frac{1}{17}$$

(for general X,Y the lower bound is $\Omega(1)\cdot T^{\frac{1}{2}}\cdot d_2(P_X;P_Y)$).

The corollary shows that the *total absolute advantage* over all partition slices, is as expected. Since $\{S_i\}_i$ is a partition we have

$$\sum_{i=1}^{T}\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \sum_{i=1}^{T}\left|\sum_{x\in S_i}\left(P_X(x)-P_Y(x)\right)\mathsf{D}(x)\right| = \sum_x\left(P_X(x)-P_Y(x)\right)\mathsf{D}(x)\beta(x)$$

where for $\beta_i \stackrel{def}{=} \text{sgn}\left(\sum_{x\in S_i}(P_X(x')-P_Y(x))\,\mathsf{D}(x)\right)$ (the sign of the advantage on the i-th slice) we define $\beta(x)=\beta_i$ where S_i contains x. This shows that by "flipping" the distinguisher output on the slices we achieve the sum of individual advantages. Since the bit $\beta(x)$ can be computed with $O(T)+\tilde{O}(n)$ advice (the complexity of the function $i\to\beta_i$ plus the complexity of finding i for a given x) we obtain

▶ **Corollary 16** (Computing total advantage by one distinguisher). *For X,Y as in Corollary 11, D and $\{S_i\}_i$ defined above there exists a modification to D which in time $\tilde{O}(n)$ and advice $O(T)$ achieves advantage $\frac{1}{3}\cdot T^{\frac{1}{2}}2^{-\frac{k}{2}}\delta$ with probability $\frac{1}{17}$.*

Finally by setting $\epsilon = T^{\frac{1}{2}}2^{-\frac{k}{2}}\delta$ and manipulating T we arrive at

▶ **Corollary 17** (Continue tradeoff). *For any ϵ there exists T such that the distinguisher in Corollary 16 has advantage ϵ and circuit complexity $s=O\left(2^k\epsilon^2\delta^{-2}\right)$.*

4 Omitted Proofs

4.1 Proof of Lemma 5 (Strengthening of Marcinkiewicz-Zygmund's Inequality for $p=4$)

Let $Z=\sum_x \xi(x)$. Since $\xi(x)$ are (in particular) 2-wise independent with zero mean, we get

$$\mathbb{E}\left(\sum_x \xi(x)\right)^2 = \sum_{x,y}\mathbb{E}\left(\xi(x)\xi(y)\right) = \sum_{x=y}\mathbb{E}\left(\xi(x)\xi(y)\right) = \sum_x \text{Var}(\xi(x))\,,$$

(the summation taken over $x,y\in\mathcal{X}$). The fourth moment is somewhat more complicated

$$\begin{aligned}
\mathbb{E}\left(\sum_x \xi(x)\right)^4 &= \sum_{x_1,x_2,x_3,x_4}\mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right)\\
&= \sum_{x_1=x_2=x_3=x_4}\mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right)+\\
&\quad +3\sum_{x_1=x_2\neq x_3=x_4}\mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right)\\
&= \sum_x \mathbb{E}\,\xi(x)^4 + 3\sum_{x\neq y}\mathbb{E}\,\xi(x)^2\,\mathbb{E}\,\xi(y)^2\\
&= 3\left(\sum_x \mathbb{E}\,\xi(x)^2\right)^2 - 2\sum_x \mathbb{E}\,\xi(x)^4\,.
\end{aligned}$$

The second equality follows because whenever $\xi(x)$ occurs in an odd power, for example $x=x_1\neq x_2=x_3=x_4$, the expectation is zero (this way one can simplify and bound also higher moments, see [7]). It remains to estimate the first moment. By Corollary 4 and bounds on the second and fourth moment we have just computed we obtain

$$\frac{1}{\sqrt{3}}\cdot\left(\sum_{x\in\mathcal{X}}\text{Var}(\xi(x))\right)^{\frac{1}{2}} \leqslant \mathbb{E}\left|\sum_{x\in\mathcal{X}}\xi(x)\right|$$

and the upper bound follows by Jensen's Inequality (with constant 1).

4.2 Proof of Lemma 8 (Characterizing smooth min-entropy)

Suppose that $\mathsf{H}_\infty^\delta(X) \geqslant k$. then, by definition, there is Y such that $\mathsf{H}_\infty(Y) \geqslant k$ and $\sum_{x:P_X(x)>P_Y(x)} P_X(x) - P_Y(x) \leqslant \delta$. Since all the summands are positive and since $P_Y(x) \leqslant 2^{-k}$, ignoring those x for which $P_Y(x) < 2^{-k}$ yields

$$\sum_{x:P_X(x)>2^{-k}} P_X(x) - P_Y(x) \leqslant \delta.$$

Again, since $P_Y(x) \leqslant 2^{-k}$ we obtain

$$\sum_{x:P_X(x)>2^{-k}} P_X(x) - 2^{-k} \leqslant \delta,$$

which finishes the proof of the "$\Longrightarrow$" part.

Assume now that $\delta' = \sum_{x\in\mathcal{X}} \max\left(P_X(x) - 2^{-k}, 0\right) \leqslant \delta$. Note that

$$\sum_{x\in\mathcal{X}} \max\left(P_X(x) - \frac{1}{2^k}, 0\right) + \sum_{x\in\mathcal{X}} \max\left(\frac{1}{2^k} - P_X(x), 0\right) = \\ = 2\sum_{x\in\mathcal{X}} \left|P_X(x) - \frac{1}{2^k}\right| \geqslant 2\sum_{x\in\mathcal{X}} \max\left(P_X(x) - \frac{1}{2^k}, 0\right)$$

and therefore we have $\sum_{x\in\mathcal{X}} \max\left(2^{-k} - P_X(x), 0\right) \geqslant \delta'$. By this observation we can construct a distribution Y by shifting δ' of the mass of P_X from the set $S^- = \{x : P_X(x) > 2^{-k}\}$ to the set $\{x : 2^{-k} \geqslant P_X(x)\}$ in such a way that we have $P_Y(x) \leqslant 2^{-k}$ for all x. Thus $\mathsf{H}_\infty(Y) \geqslant k$ and since a δ' fraction of the mass is shifted and redistributed we have $d_1(X;Y) \leqslant \delta'$. This finishes the proof of the "$\Longleftarrow$" part.

4.3 Proof of Corollary 14 ((Mixed) moments of slice advantages)

For shortness denote $\Delta(x) = P_X(x) - P_Y(x)$ and $\mathsf{Adv}_{S_i}^{\mathsf{D}} = \mathsf{Adv}_{S_i}^{\mathsf{D}}(X;Y)$.

Note that by Lemma 5, applied to the family $f_x = \Delta(x)\mathsf{D}(x)\mathbf{1}_{S_i}(x)$ (which is 4-wise independent) we have

$$\mathbb{E}\,\mathsf{Adv}_{S_i}^{\mathsf{D}} \geqslant 3^{-\frac{1}{2}} \left(\sum_x \Delta(x)^2\right)^{\frac{1}{2}}$$

which is the first inequality claimed in the corollary. In turn, again by Lemma 5, we have

$$\mathbb{E}\left(\mathsf{Adv}_{S_i}^{\mathsf{D}}\right)^2 = T^{-1} \cdot \sum_x \Delta(x)^2.$$

Since this holds for any i, by Cauchy-Schwarz we get for any i, j

$$\mathbb{E}\,\mathsf{Adv}_{S_i}^{\mathsf{D}}\mathsf{Adv}_{S_j}^{\mathsf{D}} \leqslant \sqrt{\mathbb{E}\left(\mathsf{Adv}_{S_i}^{\mathsf{D}}\right)^2 \cdot \mathbb{E}\left(\mathsf{Adv}_{S_j}^{\mathsf{D}}\right)^2} \leqslant T^{-1} \cdot \sum_x \Delta(x)^2$$

which proves the second inequality in the corollary.

References

1 Bonnie Berger. The fourth moment method. *SIAM J. Comput.*, 26(4):1188–1207, 1997. doi:10.1137/S0097539792240005.

2 Anindya De, Luca Trevisan, and Madhur Tulsiani. Time Space Tradeoffs for Attacks against One-Way Functions and PRGs. In *Advances in Cryptology – CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, pages 649–665, 2010. doi:10.1007/978-3-642-14623-7_35.

3 Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs. Key derivation without entropy waste. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology – EUROCRYPT 2014*, volume 8441 of *Lecture Notes in Computer Science*, pages 93–110. Springer Berlin Heidelberg, 2014. doi:10.1007/978-3-642-55220-5_6.

4 Aleksandr Khintchine. Über einen Satz der Wahrscheinlichkeitsrechnung. *Fundamenta Mathematicae*, 6(1):9–20, 1924. URL: http://eudml.org/doc/214283.

5 J. Marcinkiewicz and A. Zygmund. Quelques théorèmes sur les fonctions indépendantes. *Studia Mathematica*, 7(1):104–120, 1938. URL: http://eudml.org/doc/218615.

6 Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *Advances in Cryptology – ASIACRYPT 2005, 11th International Conference on the Theory and Application of Cryptology and Information Security, Chennai, India, December 4-8, 2005, Proceedings*, pages 199–216, 2005. doi:10.1007/11593447_11.

7 Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-hoeffding bounds for applications with limited independence. In *Proceedings of the Fourth Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms, 25-27 January 1993, Austin, Texas.*, pages 331–340, 1993. URL: http://dl.acm.org/citation.cfm?id=313559.313797.